JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

pubs.acs.org/JAFC

Article

Distinguishing Infested Flour from Uninfested Flour through Chemometric Processing of DART-HRMS Data—Revealing the Presence of *Tribolium castaneum*, the Red Flour Beetle

Amy M. Osborne, Samira Beyramysoltan, and Rabi Ann Musah*

Cite This: *J. Agric. Food Chem.* 2023, 71, 8613−8621

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Insect infestation of agricultural stored products is a significant challenge to food security across the globe. One common pest is *Tribolium castaneum* (red flour beetle). In a new approach to addressing the threat of these beetles, Direct Analysis in Real Time-High-Resolution Mass Spectrometry was used to examine infested and uninfested flour samples. These samples were then distinguished through statistical analysis techniques, including EDR-MCR, in order to highlight the important m/z values contributing to the differences in the flour profiles. A subset of these values responsible for the identification of infested flour (nominal m/z 135, 136, 137, 163, 211, 279, 280, 283, 295, 297, and 338) were further investigated, and compounds responsible for these masses included 2-(2-ethoxyethoxy)ethanol, 2-ethyl-1,4-benzoquinone, palmitic acid, linolenic acid and oleic acid. These results have the potential to lead to a rapid technique by which flour and other grains can be tested for insect infestation.

KEYWORDS: *stored-product entomology, statistical analysis, Tribolium castaneum, postharvest grains*

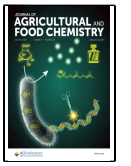
1. INTRODUCTION

Insect infestation of postharvest agricultural products such as food grains is a major challenge to food security across the globe. Infestations can be difficult to detect due to the large volume of the food stuff in comparison with the small sizes of the insects.[1] Additionally, the insects can cause further damage through the transmission of harmful micro-organisms such as fungi and bacteria as well as other toxins.[2,3] This can lead to estimated food losses of approximately 10−20% in the United States and as much as 50% in some developing countries.[2] Financial strain due to insect infestation includes not only loss of profits from contaminated products but also the expense associated with infestation monitoring and pest management. Additionally, damage to brand name reputation and the possibility of litigation if contaminated products make their way to the market are possible.[4]

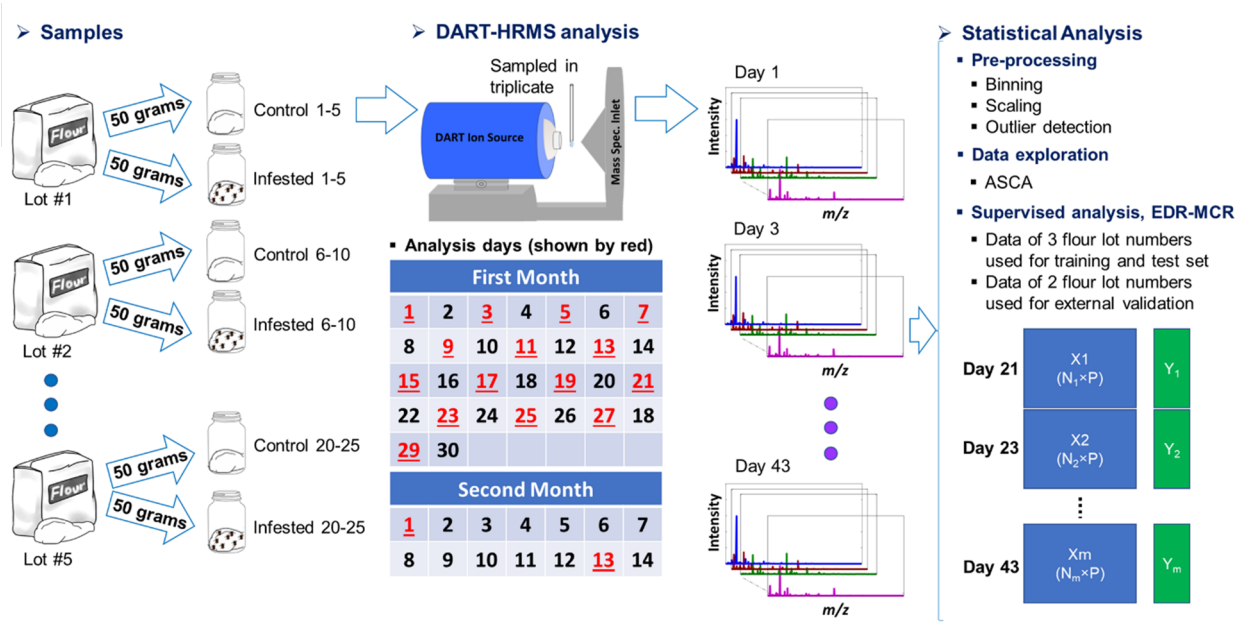
Due to shared global concerns regarding pest infestation of agricultural products, there have been a number of studies aimed at devising suitable approaches for the rapid, accurate, and sensitive detection of insect pest infestation. While many traditional methods such as traps,[5−7] microscopy analysis,[8,9] and X-ray detection[10−12] focus on determination of the presence of partial or whole insects, a recent trend in modern approaches to insect pest detection involves testing for trace volatile organic compounds (VOCs) that are produced by the insects.[13−16] Some, such as sex pheromones, are insect-species specific.[17] One study that utilized *Tribolium castaneum* determined that the most notable compounds found in the headspace of infested wheat flour were methyl-1,4-benzoquinone (MBQ), ethyl-1,4-benzoquinone (EBQ) and 1-tridecene. The quinones MBQ and EBQ are proposed defense compounds, while 1-tridecene is a male sex pheromone characteristic of members of the Tenebrionidae family.[15] A small number of recent publications have employed solid phase microextraction (SPME) fibers to concentrate the insect volatiles in the headspace of infested grain, followed by GC-MS analysis to deduce the identities of these volatiles.[18−21] While this approach has revealed the presence of insect volatiles (including dimethyltridecanal (tribolure) and 1-pentadecene as specific indicators of *T. castaneum*[18]), the large number of compounds observed in the majority of these studies make it a daunting task to identify which subset of molecules can be used to differentiate between infested and uninfested grain samples. The abundant presence of grain-derived volatiles further complicates the analysis.

One approach that may be taken to circumvent consideration of molecules that are not markers of infestation is to identify the subset that are contributed by the insect pests themselves, as this will pave the way for the development of tests for infestation that are based on the presence of specific molecules. In this regard, multivariate statistical processing of insect volatiles' signatures, with the goal of developing a classification system for distinguishing between infested and uninfested grain, could be very helpful. Pattern recognition algorithms applied to mass spectral profiles in particular could reveal the m/z values most heavily weighted in enabling the prediction model to correctly classify grain as infested or uninfested, without necessarily knowing the identities of the

Received: February 3, 2023
Revised: April 14, 2023
Accepted: April 17, 2023
Published: May 23, 2023



Scheme 1. Sample Collection and Analysis Workflow



molecules that enable the differentiation of the two types of samples.

An optimal approach for surveying harvested grains requires the rapid collection of small-molecule profile data so that pest management plans can be implemented swiftly. To this end, Direct Analysis in Real Time−High-Resolution Mass Spectrometry (DART-HRMS) may be a useful technique for high-throughput analysis, because it could rapidly reveal the chemical signatures that are characteristic of infested and uninfested grains.[22] These data could then be processed by machine learning pattern recognition tools to facilitate development of a sample analysis workflow for assessment of infestation. Furthermore, a technique that enables the ability to directly examine samples of postharvest products in their native form would be an added bonus, as it would eliminate hazardous waste produced from lengthy extractions and promote a green chemistry approach to the analysis. In this regard the recently developed Efficient Data Reduction−Multivariate Curve Resolution (EDR-MCR) method[23] facilitates classification of the type of complex data generated by reducing the data dimensionality prior to the classification. Thus, using this data processing method, the infested grain can be detected and distinguished from uninfested product in a simple and rapid fashion.

In order to develop such an approach, a suitable system comprised of a food commodity that is susceptible to infestation along with an insect pest is needed. In this work, milled flour and the red flour beetle, *Tribolium castaneum*, were used. *T. castaneum* is a stored product pest prevalent around the world due to its ability to survive in diverse climates.[24] Despite its moniker, it infests a wide range of products including grains, cereals, beans, spices, seeds, etc.[19] The most common techniques used for detecting red flour beetle infestation in grain include the use of insect traps/funnels and flotation.[25] However, these methods are primarily only useful for identifying adult insects, are not species specific, and it is possible to miss the presence of beetles if the particular aliquot of infested material that is sampled does not happen to have insects present within it. It would be beneficial to develop a technique that can be more broadly applied to not only detect multiple insect life stages, but also will detect infestation independent of whether the segment of material sampled actually has insects within it. Reported here is the development of a novel and rapid approach that utilizes chemometric processing of DART-HRMS-derived chemical signatures for the accurate detection of *T. castaneum* infestation in milled flour, based on detection of insect-derived chemical traces.

2. MATERIALS AND METHODS

2.1. Instrumentation. A DART SVP ion source (IonSense, Saugus, MA, USA) coupled to a JEOL AccuTOF mass spectrometer (JEOL USA, Peabody, MA, USA) was used. Analyses were performed in positive ion mode, and calibration was conducted with polyethylene glycol (PEG) as a standard for accurate mass determinations. Orifice 1 was set to 20 V for soft ionization; orifice 2 and the ring lens were set to 5 V, and the ion guide was set to 400 V. Ultra-high-purity helium (Airgas, Albany, NY, USA) was used at a flow rate of 2 L min^{-1}. The DART gas stream temperature was set to 350 °C, and the gap width was set to 1 cm. Spectra were collected at a rate of 1/s. Spectral data calibration, spectral averaging, and background subtraction were accomplished using TssPro3 software (Shrader Software Solutions, Detroit, MI, USA). Visual examination of the data was facilitated using Mass Mountaineer software (RBC Software, Portsmouth, NH, USA).

An Agilent 7890B gas chromatograph coupled to an Agilent 5977B mass spectrometer (Agilent, Santa Clara, CA, USA) and GERSTEL multipurpose sampler (MPS) (GERSTEL; Linthicum, MD, USA), running on Qualitative Navigator software (Agilent, Santa Clara, CA, USA) was used for GC-MS analyses. The instrument was equipped with a split/splitless injector and a DB5MS UI column (30 m, 0.25 mm). A secondary Agilent 7890B gas chromatograph coupled to a JMS Q1500 quadrupole mass spectrometer (JEOL, Peabody, MA, USA) and HTA GC Autosampler (HTA, Brescia, Italy) was also utilized. Both instruments were operated using similar parameters. Ultra-high-purity helium gas (Airgas, Albany, NY, USA) was used at a constant flow rate of 1 mL/min. The instruments were operated in splitless mode using an inlet temperature of 250 °C and sample injection volume of 1 μL. The initial oven temperature was set to 60 °C, which was held for 1 min before increasing at a rate of 10 °C/min to 100 °C. The temperature was then increased at a rate of 5 °C/min to 300 °C, where it was held for 15 min. For the mass spectrometer,

https://doi.org/10.1021/acs.jafc.3c00685
J. Agric. Food Chem. 2023, 71, 8613−8621

the ionization mode was EI, the ion source temperature was 230 °C, the m/z range was 35−1000, and there was a 2 min solvent delay. Data processing was accomplished using MassHunter qualitative analysis software (Agilent, Santa Clara, CA, USA) and AnalyzerPro XD software (Spectralworks Limited, Runcorn, UK).

2.2. Sample Acquisition and Set Up. Gold Medal brand all-purpose bleached flour representing differing lots was purchased from Walmart (Albany, NY, USA), Market 32 (Newburgh, NY, USA), Shop Rite (Niskayuna, NY, USA), Whole Foods Market (Albany, NY, USA), and Hannafords (Niskayuna, NY, USA) in order to account for natural variations in chemical profiles as a result of differences in factory processing, as well as differences in flour batches. Ten units of *T. castaneum* beetles were purchased from Carolina Biological Supply Company (Burlington, NC, USA). It was unknown at the outset of this study the point at which flour would be adequately infested with insects to exhibit a sufficiently different chemical signature to be detectable. Therefore, a time course study was conducted to monitor the changes to the flour chemical signature, starting at the time of the initial infestation.

Flour was infested with *T. castaneum* and sampled frequently for a period of several months. In replicates of 5 for each lot number of flour purchased, 25 insects were added to approximately 50 g of flour, with samples being incubated at 29−30 °C. The tops of the jars were secured with cheesecloth to allow the insects to respire while preventing their escape. Control samples were also established in replicates of 5 for each lot of flour by placing 50 g of flour in a jar and sealing it with a metal lid. The control samples were also housed in the incubator. For each day of analysis, beginning on Day 1 when samples were set up and continuing every other day for 1 month, plus an additional sampling on Day 43, 50 mg of flour was removed from the jar for each sample. In total, 17 days of sampling occurred. Dry flour was analyzed directly. However, in order to determine the form in which the most amount of chemical profile information could be revealed, extracts were also prepared using 50 mg of flour and 1 mL of each of the following solvents: water, methanol, ethyl acetate, and hexanes. Samples were vortexed for 30 s and allowed to stand overnight. Extracts were then centrifuged at 5000*g* and filtered into vials. Dry samples and extracts were analyzed by exposing the closed end of a melting point capillary tube to the sample and presenting the coated surface of that capillary tube to the open-air space between the mass spectrometer inlet and the DART ion source. Spectra were collected at a rate of 1 spectrum per second over the mass range m/z 60−1000, with each spectrum being generated in ∼3 s. The analysis of five replicates of each of the five lots of flour resulted in 25 spectra of control samples and 25 spectra of infested samples per day for each type of sample (i.e., bulk dry vs extracts). A representation of the sample collection and analysis workflow is presented in Scheme 1—Samples/DART-HRMS Analysis. All analytical standards were obtained from Sigma-Aldrich (St. Louis, MO, USA).

2.3. Statistical Analysis. *2.3.1. Data Preprocessing and Exploration.* DART high-resolution mass spectra acquired from analysis of control and infested samples were processed using TSSPro3 software (Schrader Analytical Laboratories, Detroit, MI, USA) for spectral centroiding, background subtraction, and calibration. Then, mass spectra were stored in text format and imported into MATLAB 9.9.0, R2020b software (The MathWorks, Inc., Natick, MA, USA) for further analysis. Scheme 1—Statistical Analysis illustrates the steps taken for the statistical analysis of the mass spectral data. First, the data representing control (or uninfested samples) and infested samples (those taken from the flour deliberately infested with red flour beetles) across the multiple days of sampling were aligned along common m/z values and binned. The optimal bin width and relative intensity cutoff value were determined by iterative application of different bin widths and relative intensity cutoffs and examination of the results of supervised classification to determine which values provided optimal classification results. The mass spectral data were subjected to logarithm (log) transforming and scaled. Following this, the data set was explored using ANOVA-simultaneous component analysis (ASCA)[26] to provide insights into the variations of the chemical profiles of flour as a result of infestation. The data were then analyzed by principal component analysis (PCA) in combination with Q residuals and Hotelling's T^2 statistic values to determine the presence of outliers, which were then removed. For supervised classification of the data generated from regular sampling over the course of 43 days, multiblock methods were employed.

2.3.2. Supervised Analysis by EDR-MCR. The recently devised efficient data reduction-multivariate curve resolution (EDR-MCR)[23] method was extended for multiblock analysis of this data in order to recognize and evaluate the important markers for the discrimination of control and infested samples. Details regarding the function of the EDR-MCR method have been published.[23] Briefly, the EDR-MCR approach is comprised of two main parts: EDR, based on PCA and convex geometry, was responsible for data splitting and variable selection or both; and MCR, an extension of multivariate curve resolution-alternative least-squares (MCR-ALS), was used for supervised classification and resolving pure response variables. As shown in Scheme 1—Statistical Analysis, EDR-MCR used column-wise augmented matrices representing spectra acquired at different time periods on different sampling days, for multiblock analysis. The matrices were stacked one on top of the other to retain the same number of columns. For the MCR run, the infested and control labels of the training samples were included as a hard equality constraint within the MCR platform. The EDR-MCR resulted in class and variable weight profiles in which each weight profile defined the weight or importance of each variable in identifying its corresponding class profile.

3. RESULTS/DISCUSSION

3.1. DART-HRMS Results. An initial visual inspection was conducted to assess the spectra generated from the solvent samples in comparison with the dry flour samples in order to determine which would provide the broadest range of m/z values reflective of the small-molecule chemical composition of the samples. Figure 1 shows representative results of the DART-HRMS analyses of infested flour in comparison with uninfested flour in positive ion mode, rendered as head-to-tail plots, with the top spectrum (blue) and the bottom spectrum (red) being infested and uninfested samples, respectively. Panels A, B, C, D, and E illustrate the results for dry flour, and water, methanol, ethyl acetate, and hexanes extracts, respectively. In all cases, the samples from uninfested spectra contained fewer peaks than those from infested spectra. This was not unexpected, since it was anticipated that the insects would introduce new molecules into the flour matrix. The extracts of uninfested and infested flour contained a number of the same prominent peaks, a consequence of their having been derived from a common matrix (i.e., flour). The water extract uninfested and infested samples (panel B) were found to contain 106−211 peaks above a relative abundance threshold of 1%. Prominent peaks shared between these spectra included nominal m/z 85, 97, 104, 127, 145, 163, 180, 198, 268, 289, 342, and 371. The methanol (panel C), ethyl acetate (panel D) and hexanes (panel E) uninfested and infested spectra were found to contain 165−178, 169−220, and 182−314 peaks above a relative abundance threshold of 1%, respectively. The methanol extract samples shared prominent peaks at nominal m/z 89, 124, 135, 145, 163, 180, 198, 254, 263, 282, 298, 355, 371, 397, 531, 563, 575, and 617. The ethyl acetate extracts shared prominent peaks at nominal m/z 89, 135, 177, 263, 281, 298, 371, 397, 451, 531, 575, and 634. Finally, the hexanes extract spectra shared prominent peaks at m/z 89, 101, 109, 135, 157, 171, 207, 263, 281, 298, 313, 355, 371, 397, 411, 451, 475, 484, 531, 548, 575, 617, and 634.

Figure 1 (panel A) also shows representative DART mass spectra of *T. castaneum* infested and uninfested bulk flour (top

https://doi.org/10.1021/acs.jafc.3c00685
J. Agric. Food Chem. 2023, 71, 8613−8621

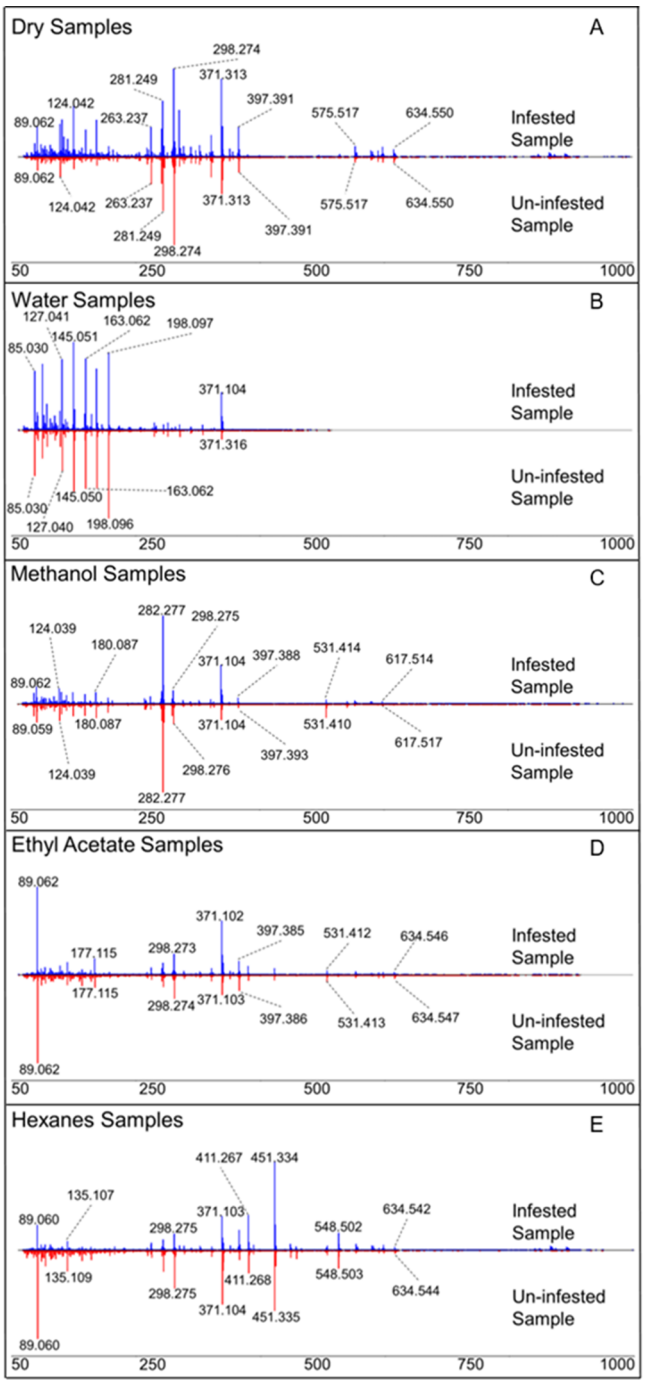


Figure 1. Representative spectra from the flour analyses arranged in head-to-tail plots comparing the infested samples (top) to the uninfested samples (bottom). From top to bottom, the panels represent: (A) bulk dry flour samples; (B) water extracted samples; (C) methanol extracted samples; (D) ethyl acetate extracted samples; and (E) hexanes extracted samples.

and bottom spectra, respectively) in positive ion mode. The bulk dry sample (panel A) spectra had 392 peaks in the uninfested spectrum and 394 peaks in the infested spectrum above a threshold of 1%. Like the extracts, the bulk dry infested and uninfested samples also shared a number of prominent peaks such as nominal m/z 89, 124, 130, 145, 163, 171, 180, 198, 205, 211, 245, 257, 263, 279, 281, 298, 306, 313, 337, 355, 371, 383, 397, 551, 575, 599, 610, 617, and 634. As the representative dry flour spectra consistently revealed the greatest number of peaks when compared to the extracts, it was concluded that the extraction step reduces the number and diversity of compounds observed in both the uninfested and infested flour. Additionally, most of the prominent peaks observed in the extract spectra were present in the bulk dry flour spectra. It was thus concluded that it was suboptimal to utilize solvent extracts, and sufficient as well as preferable to analyze the bulk dry sample in order to glean the greatest amount of chemical information about potential markers that would enable infested and uninfested samples to be differentiated. This is advantageous, because it indicates that the performance of an extraction step may not be required to detect infestation markers of interest, and this would enhance the speed and efficiency of the analysis procedure.

3.2. Statistical Analysis of the Data. A number of statistical analysis techniques were applied to the data in order to determine how best to differentiate the infested and uninfested flour and also determine which m/z values might be most important for enabling the two sample types to be distinguished. To this end, the mass spectral data representing control and infested samples collected over the course of 43 days were binned along common m/z values. A peak corresponding to a plasticizer derived from the capillary tubes used for sampling (nominal m/z 371) was removed from the mass spectra prior to statistical analysis processing of the data. The optimal bin width and relative abundance threshold cutoff were found to be 15 mmu and 1%, respectively. This treatment resulted in matrices with dimensions of 50 × 1191 for each storage day, representing 25 spectra for each control and infested flour, composed of 5 replicates of 5 flour lots (illustrated in Figure S1).

3.2.1. Data Exploration. The log transformed data were mean-centered and then subjected to ASCA to reveal the effect of red flour beetle infestation and the length of storage on the chemical profiles of the flour and to reveal common patterns across the different days. ASCA decomposes the multivariate data set into factor effect matrices, in which each effect matrix includes the variability induced in the data by an experimental factor (factors here include the infestation by the red flour beetles and the length of storage represented by the day the sample was analyzed). The effect value of each experimental factor was calculated as the sum of the square of the effect matrix values, which quantify the factor variation underlying the data set. In order to validate the ASCA results, a permutation test was performed to quantify the significance of the experimental factors using computed p-values. The labels of samples for data associated with days 3−43 were shuffled randomly in each iteration, and then ASCA was implemented on the shuffled data to compute the effect values corresponding to experimental factors. From this treatment, the distributions of the computed effect values could be determined. Figure 2A−B presents histogram plots showing the effect value distributions resulting from the permutation test (200 iterations) of ASCA for the factors of insect infestation and storage length, respectively. The red arrows in the plots illustrate the effect value of the original groupings (labels were not permuted); if the red arrow is on the right side and outside of the distribution (shown by the gray bars), the original effect value is significantly different from the permuted effect values. In Figure 2A it is shown that the red arrow is superimposed on the distribution observed by permuting the data, and the approximate p-value is 0.765 (>0.005). This means that the variability of the chemical profiles that may be

https://doi.org/10.1021/acs.jafc.3c00685
J. Agric. Food Chem. 2023, 71, 8613−8621

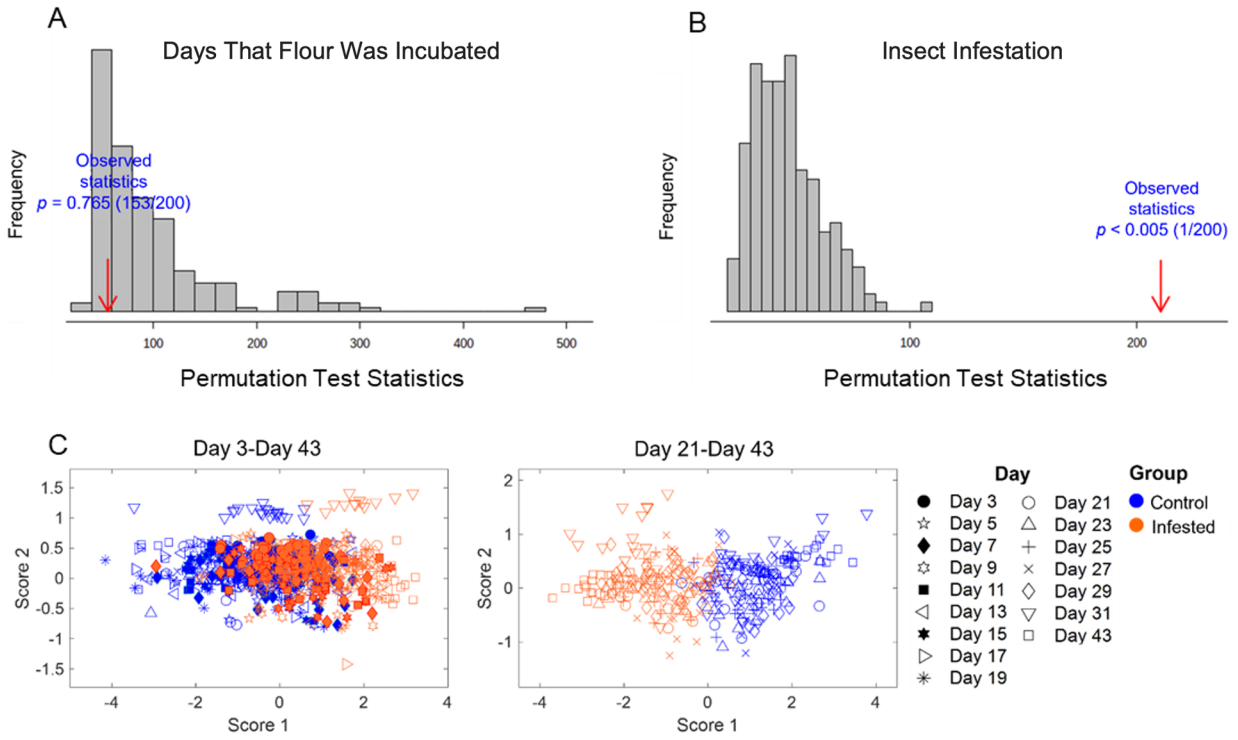


Figure 2. Determination of the significance of the variables *time* and *infestation status* using ASCA and permutation tests. Panels A and B are histogram plots of permutation test results of ASCA applied to day 3−43 data, to illustrate the significance of the influence of length of insect infestation on the chemical profile of the flour. The permutation test result (in 200 iterations) revealed that the impact of infestation on the flour chemical profile is significant when considering a *p*-value threshold of 0.005. However, the impact of time (reflected by the sample collection day) on the flour chemical profile was not significant. Panel C is the score plot derived from ASCA for samples analyzed on days 3−43 (specifically, days 3, 5, 7, 9, 11, 13, 15, 17, 21, 23, 25, 27, 29, 31, 35 and 43) and days 21−43 (specifically days 21, 23, 25, 27, 29, 31, 35 and 43). The blue and orange points illustrate the samples related to control and infested flours, respectively.

the result of storage length is not significant at a threshold value of 0.005. In contrast, in Figure 2B, the effect values for all permuted data (the gray bars) are smaller than the effect value of the original groupings (shown with red arrows) which means that the effect of infestation on flour chemical profiles is significant (the approximate *p*-value is lower than 0.005). In summary, what panels A and B reveal is that it is the presence of the insects over time, and not the number of days that the flour has been incubated, that results in a statistically significant change in chemical profile. Figure 2C displays the results of ASCA implementation on data collected on days 3−43 and days 21−43. The score plots in the figure illustrate the projections of the data on the first two principal components (Score 1 − Score 2) of the insect infestation matrix and reveal a pattern of discrimination between infested and uninfested flour. The samples collected on different days are illustrated by different markers in the plot (blue and orange colors show the control and infested samples, respectively). While the permutation test[27] (Figure 2B) performed on the effect matrices revealed that the effect of infestation on the flour chemical profiles is significant, the score plot in Figure 2C−Day 3−Day 43 does not illustrate a clear separation between the control and infested flour beyond a tendency of the control samples to lean to the left of the plot and a tendency of the infested samples to lean to the right. However, exploration with ASCA for days 21−43 (Figure 2C−Day 21−Day 43) revealed clustering of control and infested samples and separation between the classes. This apparent clustering could be due to the presence of insect-derived markers and/or a change in the concentration of flour compounds over the course of the infestation. What Figure 2C reveals is that it is at the ∼3-week time point that the DART-HRMS chemical profile of infested flour is distinct enough to show clear separation between the clusters in the score plot corresponding to days 21−43.

3.2.2. Supervised Analysis. Accordingly, data collected after day 19 (i.e., days 21, 23, 25, 27, 29, 31, and 43) were further analyzed using supervised methods in order to determine the variables most important in enabling the model to differentiate between infested and control samples. To create a model for detection of infestation and reveal the important *m*/*z* values, the data collected from three flour lots were utilized for the training step, and the data from two other lots were used for the external validation step. Since EDR is sensitive to outliers, the PCA components of the flour data that explained 98% of the data variance were examined for the presence of sample outliers, which were identified and removed, leaving 900 samples remaining for the analysis. Then, each spectrum was normalized to the base peak, and after log transformation, the data were subjected to EDR-MCR to provide class and variable weight profiles. Accordingly, EDR was applied to the data set corresponding to each collection day in order to define the training and test sets and reduce the number of variables. The explained variance threshold in the EDR algorithm was set to 99%. EDR reduced the number of *m*/*z* values considered from 1192 to 52, which are listed in Table S1. To treat the multiple matrices representing the different sampling days, MCR was applied for the classification of this multiblock data. MCR-ALS[28] (which is a component of the EDR-MCR algorithm) is well-known for its ability to analyze multiple data matrices in a

https://doi.org/10.1021/acs.jafc.3c00685
J. Agric. Food Chem. 2023, 71, 8613−8621

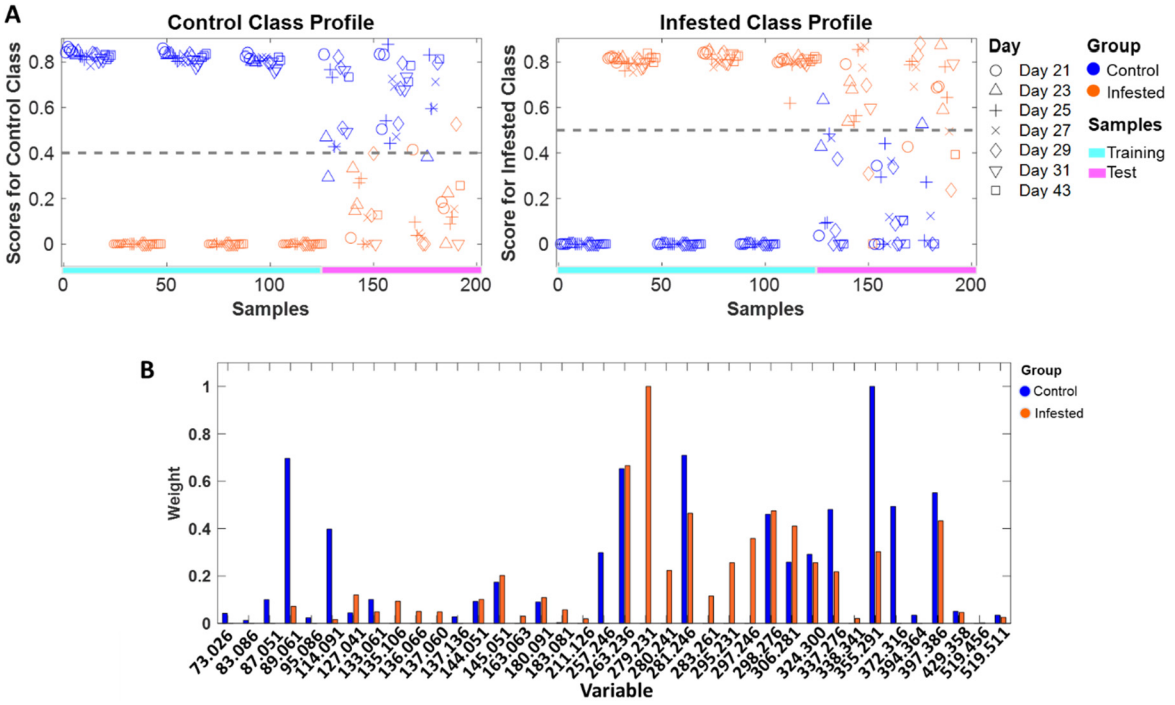


Figure 3. Results of the application of EDR-MCR to the analysis of the flour data set. Class profiles relevant to control and infested flour, resolved by EDR-MCR, are shown in panel A. The dashed lines in the plots display the thresholds (at 0.4 and 0.5 for control and infested class profiles, respectively) that were used for class assignment. (B) Scaled weight profiles resulting from EDR-MCR demonstrate the relative differences between the weights of each m/z value (shown on the x-axis) for infested and control classes.

column- or a row-wise multiset configuration. In a column-wise arrangement, it is assumed that data matrices share their column vector space which is a condition fulfilled for data sets from different sample days, as they share all or some of the same m/z values. Conversely, the row vector space of each data matrix is left unshared and independent of the other matrices. The MCR algorithm of the EDR-MCR was applied to the data with 8 MCR components, and a non-negativity constraint was imposed on all profiles. Labels of "control" and "infested" were assigned to the training samples as hard equality constraints in the MCR process. EDR-MCR resulted in two matrices representing the class and variable weight profiles shown in blue and orange colors in Figure 3A and B, respectively. Figure 3A shows the results of data analysis, with the right and left panel plots displaying the class profiles for the control and infested classes, respectively. The samples belonging to different days are indicated with different marker shapes, and additionally, the training and test samples are indicated with cyan and magenta colors, respectively, on the x-axes of the plots. The gray dashed lines in the plots indicate the class score thresholds for class assignment (0.4 and 0.5 for control and infested class profiles, respectively). Evaluation results (based on accuracy, sensitivity, specificity, and precision) determined using test and external validation samples are presented in Table S2. According to the prediction results for the test samples, the sensitivity and specificity for infestation recognition are 0.83 and 0.76, respectively. The external validation results showed the sensitivity and specificity of infestation identification to be 0.79 and 0.86, respectively. Relying on the validation results, the EDR-MCR model was well-enough fitted to discriminate the control from the infested samples. The markers most heavily weighted in enabling discrimination between the two classes could then be extracted. Accordingly, the weight profiles resulting from EDR-MCR were scaled to show the relative differences between the weights of each variable for each class. From this treatment, a subset of the 52 variables that were revealed by EDR-MCR to have a weight of zero for discrimination could be removed from the bar plot, leaving 37 variables that were identified as being important for the discrimination of infested flour and control (uninfested) flour. These m/z values and their weights are presented in Figure 3B. In the figure, the blue bars show the variables associated with the control samples and their weights, and the orange bars are representative of the weights of the variables for the infested samples. While some of the variables are important for classification of both infested and uninfested flours (e.g., m/z 263.2384), other markers were important to infested or uninfested flour. These included m/z 279.2341, 280.2417, 283.2622, 295.2304 and 297.2478 among others, which have importance for distinguishing infested flour and zero weight for the uninfested flour; and m/z 87.0491, 137.1373, 257.2476, 372.3183 and 394.3603 among others, which have importance for uninfested flour and zero weight for distinguishing infested flour. It should be mentioned that m/z 137.0501 was detected as the only marker which was found only in spectral data collected of infested flour. In summary, the Figure 3B plot shows the masses most heavily weighted in the differentiation of infested and control flour. Following the discovery of the subset of m/z values indicative of infestation, the process of their structure determination was initiated.

3.3. Compound Identification. Of the 37 variables shown to be important in differentiating between infested and uninfested flour, a subset of 11 of them were important for the classification of infested flour. These were m/z 135.1062, 136.0661, 137.0501, 163.0633, 211.1262, 279.2341, 280.2417,

https://doi.org/10.1021/acs.jafc.3c00685
J. Agric. Food Chem. 2023, 71, 8613−8621

Table 1. Structure Confirmation of Compounds Determined to Be Important for Distinguishing between *T. castaneum* Infested and Uninfested Flour, Based on Retention Time and EI-MS Fragmentation Pattern Matching with Authentic Standards

Compound	Formula	Standard RT (min)	Sample RT (min)	Solvent
2-(2-Ethoxyethoxy)ethanol	$C_6H_{14}O_3$	5.6	6.09	Methanol
Palmitic Acid	$C_{16}H_{32}O_2$	27.8	27.1	Ethyl Acetate
Linoleic Acid	$C_{18}H_{32}O_2$	29.5	30.0	Ethyl Acetate
Oleic Acid	$C_{18}H_{34}O_2$	30.6	30.7	Ethyl Acetate [MRA1]

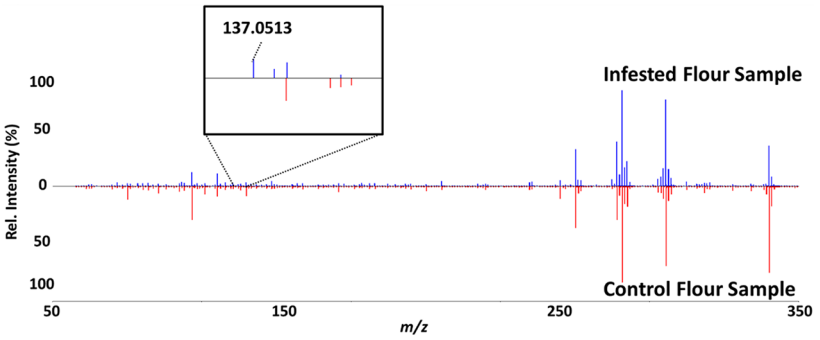


Figure 4. Head-to-tail plot displaying an infested flour sample mass spectrum on the top and a control flour sample mass spectrum on the bottom. The inset panel highlights nominal m/z 137 (EBQ) which is present in the infested sample but absent from the control sample.

283.2622, 295.2304, 297.2478, and 338.3411. An attempt was made to determine the identities of these molecules. Ethyl acetate and methanol extracts of the infested flour were analyzed by GC-MS to determine whether any of the observed peaks exhibited molecular weights of any of the 11 infested flour specific masses. Of the peaks observed in the methanol extract flour sample, those appearing at retention times 5.6 and 29.7 min revealed mass spectra of molecules with molecular weights that were consistent with those of m/z 135.1062 and 283.2622, which were found to be important primarily for differentiation of infested flour samples. Of the peaks observed in the ethyl acetate extract, those appearing at retention times 27.0, 30.3, and 30.7 min revealed mass spectra of molecules with molecular weights consistent with m/z 256.240, 280.240, and 282.256, corresponding to protonated masses of 257.2476, 281.2481, and 283.264. It should be noted that though nominal m/z 257 and 281 were not exclusively weighted in distinguishing infested flour samples, they were still deemed important to the differentiation of both the control and infested flour samples (based on the earlier described EDR-MCR analysis). An additional chromatogram of an infested flour sample extracted with ethyl acetate also exhibited a peak at 7.8 min, which was found to have a mass spectrum that aligned with m/z 136.0520 which was observed protonated in the DART high-resolution mass spectra of bulk dry infested flour samples as m/z 137.0501. Because the DART-HRMS analyses furnished high-resolution masses, corresponding molecular formulas and tentative structures could be assigned to the compounds associated with each of the retention times. These were 2-(2-ethoxyethoxy)ethanol ($C_6H_{14}O_3$), 2-ethyl-1,4-benzoquinone (EBQ, $C_8H_8O_2$), palmitic acid ($C_{16}H_{32}O_2$), linolenic acid ($C_{18}H_{30}O_2$), and oleic acid ($C_{18}H_{34}O_2$), which correspond to the protonated masses 135.1062, 137.0501, 257.2476, 279.2341, and 283.2622, respectively. The tentative assignments were subsequently confirmed for all compounds for which authentic standards were commercially available, based on GC retention time and EI-MS fragmentation pattern matching (standards were available for all the tentatively assigned compounds except EBQ). These results are reported in Table 1, and the corresponding chromatograms are shown in Figure S2. Additionally, Figure 4 shows a head-to-tail plot with an infested sample spectrum displayed on the top and a control sample spectrum displayed on the bottom, highlighting the presence of EBQ in the infested flour and its absence in the uninfested sample. Though standards were unavailable, NIST mass spectral database searches of the mass spectra corresponding to specific peaks observed in the chromatogram of infested flour matched to EBQ. The match to EBQ is significant, as it is a compound whose presence is linked specifically to red flour beetles.[3,15,29−31] Interestingly, while methyl-1,4-benzoquinone, ethyl-1,4-benzoquinone, 1-tridecene[15] tribolure and 1-pentadecene[18] are compounds previously reported to be associated specifically with *T. castaneum* infestation, only EBQ was detected in our study, and no masses corresponding to these compounds were observed in our DART-HRMS analyses of infested flour. It remains unclear whether the discrepancy is a function of the difference in the methods used in the studies or the difference in chemical profiles that resulted from the use of different milled grains (milled all-purpose flour from the United States versus whole CWRS wheat from Canada and ground wheat from Australia).

Overall, the results reveal that there is a subset of compounds whose presence in flour may be reflective of infestation because they appear important for the differentiation of infested flour. Some of these compounds, such as the linoleic and oleic acids, have not previously been reported to be associated with *T. castaneum* infestation. In principle, detection of any or a combination of these by any one of a variety of analytical methods could be used as a test for *T. castaneum* infestation. What remains unknown is the correlation between level of infestation and the amount of these compounds that are markers for infestation. The establishment of this relationship will allow determination of the approximate number of insects per unit mass of material,

https://doi.org/10.1021/acs.jafc.3c00685
J. Agric. Food Chem. 2023, 71, 8613−8621

based on the amounts and identities of trace compounds. This will be the subject of future investigations.

It is noteworthy that due to the nature of the statistical analysis applied to the data, the determination of infestation can be made in the absence of knowledge of the identities of the discriminating m/z values. While only one species and one type of grain were studied in this report, the results illustrate that the application of chemometrics and DART-HRMS to postharvest product analysis could offer significant improvement in the agricultural industry in terms of time and resources spent to investigate insect infestation. Compared to other methods of analysis for detection of the insect volatiles in agricultural products, DART-HRMS is a much more rapid technique and it can be performed by direct analysis of bulk dry flour, with no sample pretreatment steps. Through further analysis of various postharvest products and the common insect pests responsible for their contamination, it may be possible to devise a machine learning-based workflow capable of distinguishing infested grains within a matter of minutes with very little training or expertise necessary on the part of the analyst.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jafc.3c00685.

Score plots from the ASCA analysis of the flour samples on days 3−43 and 21−43 with permutation test results; chromatograms of flour extracts; m/z values extracted by EDR; and performance results of the EDR-MCR model for discrimination of infested and uninfested flour samples (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Rabi Ann Musah − *Department of Chemistry, University at Albany State University of New York, Albany, New York 12222, United States*; Phone: (518) 437-3740; Email: rmusah@albany.edu

Authors

Amy M. Osborne − *Department of Chemistry, University at Albany State University of New York, Albany, New York 12222, United States*

Samira Beyramysoltan − *Department of Chemistry, University at Albany State University of New York, Albany, New York 12222, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jafc.3c00685

Funding

The funding support of the National Institute of Justice (NIJ), Office of Justice Programs, U.S. Department of Justice (DOJ) [award 2020-MU-MU0016], as well as the University at Albany—State University of New York is gratefully acknowledged.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Thanks are extended to Dr. Robert Cody (JEOL USA Inc.) for technical assistance. The opinions, findings, and conclusions or recommendations expressed are those of the of the authors and do not necessarily reflect those of the DOJ. Parts of this work appear in the Ph.D. thesis of A.M.O.

■ ABBREVIATIONS USED

DART-HRMS, Direct Analysis in Real Time-High Resolution Mass Spectrometry; EDR-MCR, Efficient Data Reduction−Multivariate Curve Resolution; VOC, Volatile Organic Compound; SPME, Solid Phase Microextraction; GC-MS, Gas Chromatography−Mass Spectrometry; m/z, Mass-to-Charge Ratio; PEG, Polyethylene Glycol; EI, Electron Ionization; RT, Retention Time; EBQ, 2-Ethyl-1,4-benzoquinone

■ REFERENCES

(1) Parkin, E. A. Stored Product Entomology (The Assessment and Reduction of Losses Caused by Insects to Stored Foodstuffs). *Annu. Rev. of Entomol* **1956**, *1* (1), 223−240.

(2) Gorham, J. R. *Ecology and management of food-industry pests*; FDA technical bulletin; 1991.

(3) Lis, L. B.; Bakuła, T.; Baranowski, M.; Czarnewicz, A. The carcinogenic effects of benzoquinones produced by the flour beetle. *Pol. J. Vet. Sci.* **2011**, *14* (1), 159−164.

(4) Hagstrum, D. W.; Phillips, T. W. Evolution of Stored-Product Entomology: Protecting the World Food Supply. *Annu. Rev. of Entomol* **2017**, *62* (1), 379−397.

(5) Chambers, J. Overview on Stored-Product Insect Pheromones and Food Attractants. *J. Kansas Entomol. Soc.* **1990**, *63* (4), 490−499.

(6) Hagstrum, D. W.; Flinn, P. W.; Subramanyam, B.; Keever, D. W.; Cuperus, G. W. Interpretation of Trap Catch for Detection and Estimation of Stored-Product Insect Populations. *J. Kansas Entomol. Soc.* **1990**, *63* (4), 500−505.

(7) Vick, K. W.; Mankin, R. W.; Cogburn, R. R.; Mullen, M.; Throne, J. E.; Wright, V. F.; Cline, L. D. Review of Pheromone-Baited Sticky Traps for Detection of Stored-Product Insects. *J. Kansas Entomol. Soc.* **1990**, *63* (4), 526−532.

(8) Leelaja, B. C.; Rajashekar, Y.; Rajendran, S. Detection of eggs of stored-product insects in flour with staining techniques. *J. Stored Prod Res.* **2007**, *43* (3), 206−210.

(9) Park, S. I.; Bisgin, H.; Ding, H.; Semey, H. G.; Langley, D. A.; Tong, W.; Xu, J. Species Identification of Food Contaminating Beetles by Recognizing Patterns in Microscopic Images of Elytra Fragments. *PLoS One* **2016**, *11* (6), No. e0157940.

(10) Haff, R.; Slaughter, D. C. Real-time X-ray inspection of wheat for infestation by the granary weevil, Sitophilus granarius (L.). *Trans. ASABE* **2004**, *47* (2), 531.

(11) Karunakaran, C.; Jayas, D.; White, N. Detection of infestations by Cryptolestes ferrugineus inside wheat kernels using a soft X-ray method. *Can. Biosyst. Eng.* **2004**, *46*, 7.1−7.9.

(12) Toews, M.; Pearson, T. C.; Campbell, J. F. Imaging and Automated Detection of Sitophilus oryzae (Coleoptera: Curculionidae) Pupae in Hard Red Winter Wheat. *J. Econ. Entomol.* **2006**, *99*, 583−92.

(13) Dey, N.; Bhattacharya, S. Nanomolar Level Detection of Uric Acid in Blood Serum and Pest-Infested Grain Samples by an Amphiphilic Probe. *Anal. Chem.* **2017**, *89* (19), 10376−10383.

(14) Xu, S.; Zhou, Z.; Li, K.; Jamir, S. M.; Luo, X. Recognition of the Duration and Prediction of Insect Prevalence of Stored Rough Rice Infested by the Red Flour Beetle (Tribolium castaneum Herbst) Using an Electronic Nose. *Sensors* **2017**, *17* (4), 688.

(15) Senthilkumar, T.; Jayas, D.; White, N.; Freund, M.; Shafai, C.; Thomson, D. Characterization of volatile organic compounds released by granivorous insects in stored wheat. *Journal of stored products research* **2012**, *48*, 91−96.

(16) Tanaka, F.; Magariyama, Y.; Miyanoshita, A. Volatile biomarkers for early-stage detection of insect-infested brown rice: Isopentenols and polysulfides. *Food chemistry* **2020**, *303*, 125381.

(17) Williams, H. J.; Silverstein, R. M.; Burkholder, W. E.; Khorramshahi, A. Dominicalure 1 and 2: Components of aggregation pheromone from male lesser grain borerRhyzopertha dominica (F.) (Coleoptera: Bostrichidae). *J. Chem. Ecol* **1981**, *7* (4), 759−780.
(18) Laopongsit, W.; Srzednicki, G.; Craske, J. Preliminary study of solid phase micro-extraction (SPME) as a method for detecting insect infestation in wheat grain. *J. Stored Prod Res.* **2014**, *59*, 88−95.
(19) Niu, Y.; Hardy, G.; Agarwal, M.; Hua, L.; Ren, Y. Characterization of volatiles Tribolium castaneum (H.) in flour using solid phase microextraction-gas chromatography mass spectrometry (SPME-GCMS). *Food Science and Human Wellness* **2016**, *5* (1), 24−29.
(20) Seitz, L.; Sauer, D., Volatile compounds and odors in grain sorghum infested with common storage insects. *Cereal Chem.* **1996**, *73* (6), 744−750.
(21) Villaverde, M. L.; Juárez, M. P.; Mijailovsky, S. Detection of Tribolium castaneum (Herbst) volatile defensive secretions by solid phase microextraction-capillary gas chromatography (SPME-CGC). *Journal of Stored Products Research* **2007**, *43* (4), 540−545.
(22) Musah, R. A.; Espinoza, E. O.; Cody, R. B.; Lesiak, A. D.; Christensen, E. D.; Moore, H. E.; Maleknia, S.; Drijfhout, F. P. A high throughput ambient mass spectrometric approach to species identification and classification from chemical fingerprint signatures. *Sci. Rep.-UK* **2015**, *5*, 11520.
(23) Beyramysoltan, S.; Abdollahi, H.; Musah, R. A. Workflow for the Supervised Learning of Chemical Data: Efficient Data Reduction-Multivariate Curve Resolution (EDR-MCR). *Anal. Chem.* **2021**, *93* (12), 5020−5027.
(24) Islam, W. Eco-friendly approaches for the management of red flour beetle: Tribolium castaneum (Herbst). *Science Letters* **2017**, *5* (2), 105−114.
(25) Neethirajan, S.; Karunakaran, C.; Jayas, D. S.; White, N. D. G. Detection techniques for stored-product insects in grain. *Food Control* **2007**, *18* (2), 157−162.
(26) Alinaghi, M.; Bertram, H. C.; Brunse, A.; Smilde, A. K.; Westerhuis, J. A. Common and distinct variation in data fusion of designed experimental data. *Metabolomics* **2020**, *16* (1), 2.
(27) Vis, D. J.; Westerhuis, J. A.; Smilde, A. K.; van der Greef, J. Statistical validation of megavariate effects in ASCA. *BMC Bioinformatics* **2007**, *8* (1), 322.
(28) Tauler, R.; Maeder, M.; de Juan, A. Multiset data analysis: extended multivariate curve resolution. In *Comprehensive Chemometrics*, 2nd ed.; Brown, S., Tauler, R., Walczak, B., Eds.; Elsevier, 2020; Chapter 2.15, pp 305−336.
(29) Faustini, D. L.; Burkholder, W. E. Quinone-aggregation pheromone interaction in the red flour beetle. *Anim Behav* **1987**, *35* (2), 601−603.
(30) Pappas, P. W.; Wardrop, S. M. Quantification of Benzoquinones in the Flour Beetles, Tribolium castaneum and Tribolium confusum. *Prep Biochem Biotech* **1996**, *26* (1), 53−66.
(31) Unruh, L. M.; Xu, R.; Kramer, K. J. Benzoquinone levels as a function of age and gender of the red flour beetle, Tribolium castaneum. *Insect Biochem Molec* **1998**, *28* (12), 969−977.

https://doi.org/10.1021/acs.jafc.3c00685
J. Agric. Food Chem. 2023, 71, 8613−8621